ESS-DIVE Bootcamp

October 29th, 2020 10:00 - 11:00 am PDT

Madison Burrus, Fianna O'Brien

Robert Crystal-Ornelas, Zarine Kakalia, Emily Robles, Deb Agarwal, Charuleka Varadharajan, Shreyas Cholia, Valerie Hendrix, Joan Damerow, Hesham Elbashandy, Mario Melara, Makayla Shepherd, Maegen Simmonds, Cory Snavely, and Karen Whitenack



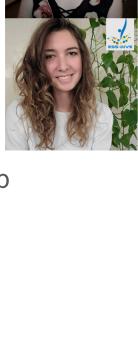












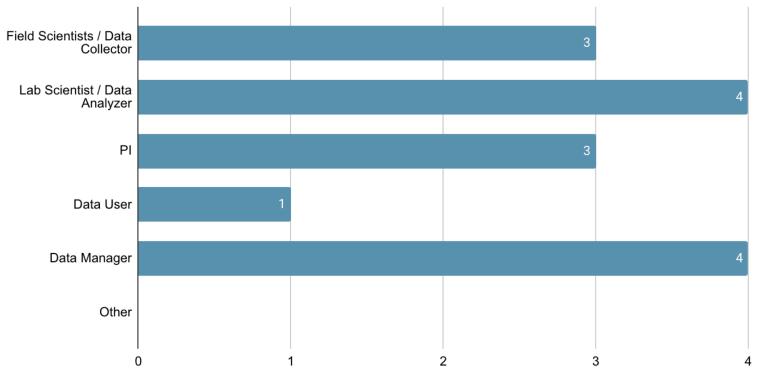




- How to register to become an ESS-DIVE data contributor
- How to submit and publish data to ESS-DIVE
- How to check the quality of your metadata before requesting to publish
- How to cite data published on ESS-DIVE

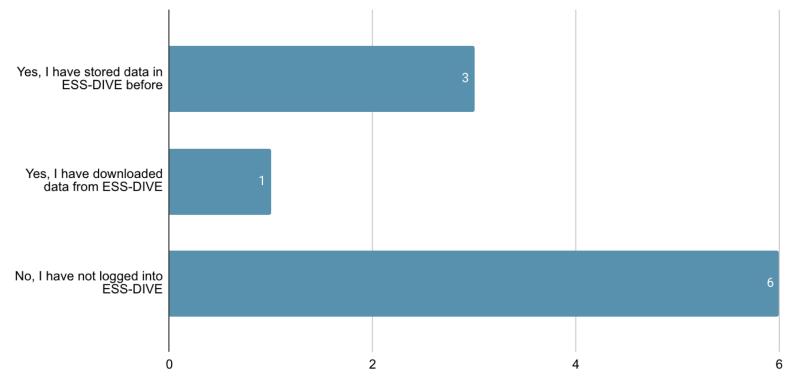
This is an interactive webinar! We have demos, polls, and question breaks prepared.

What is your primary role in your research group?

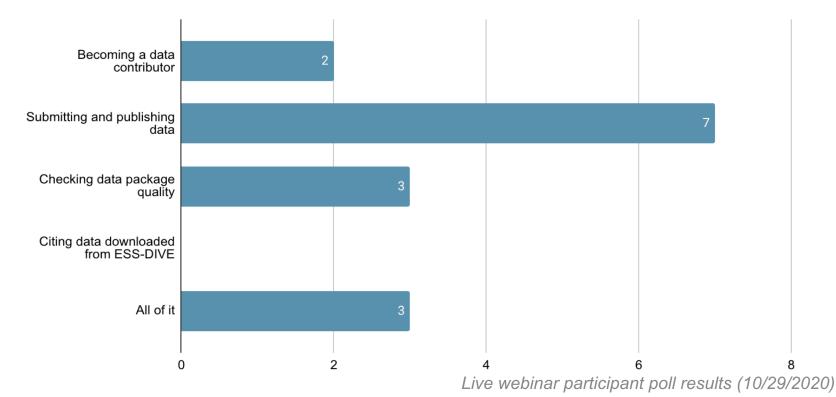






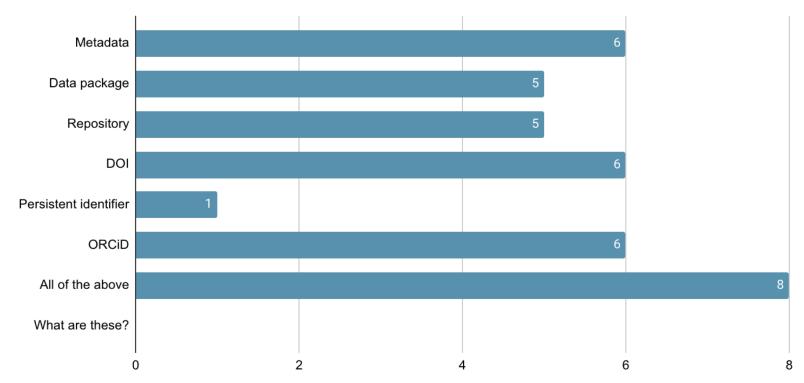


Which section are you most interested in learning about?



Which of these terms are you familiar with?





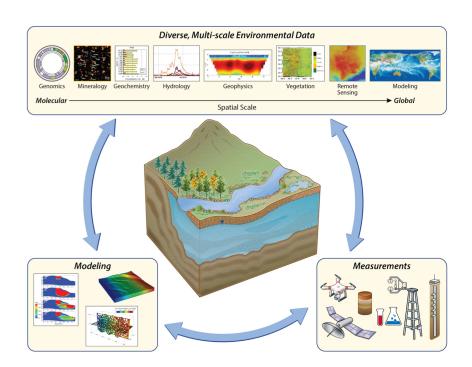


What is ESS-DIVE?

What is ESS-DIVE?



- https://ess-dive.lbl.gov/
- A Repository for Earth and Environmental Data funded by the Department of Energy & maintained by the Lawrence Berkeley National Laboratory.
- ESS-DIVE stores diverse
 environmental data types such as
 remote imagery, water quality,
 geophysical and model data



Varadharajan et al. (2019)1

ESS-DIVE's Goal



- The goal of our repository is to provide DOE-funded researchers a centralized location to store their environmental research data for longterm preservation & publication
- ESS-DIVE allows users to:
 - Upload and download data
 - Publish data with persistent identifier for citations
 - Search public data

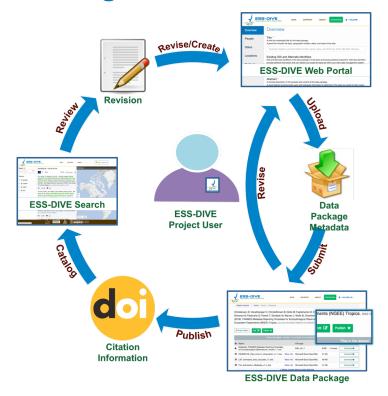






Publishing Data on ESS-DIVE





Varadharajan et al. (2019)1

It's Easy!

ESS-DIVE Support provides materials to guide users through the data publication process.

Past webinars, tutorial videos, and help documentation can be found under the Support tab on our website

Terms to keep in mind





Data Storage

Repository

 A central location for long term data storage

Published

 Data that has been reviewed and is publicly accessible



Publishing Terms

Metadata

 Descriptive information about data files

Data Package

 Term we use to refer to both data files and metadata



ORCID

 For people; permanent link for your professional activities

DOI

For data;
 permanent link on the internet



How To Become a Data Contributor





You must be approved to store data on ESS-DIVE

- Logging in to ESS-DIVE with your ORCiD lets you download data
- Registering as a data contributor lets you upload data







How to Register as a Data Contributor



Log in to ESS-DIVE



01

Using your **ORCID**, sign into **ESS-DIVE**.

Fill in Contributor Form



02

Click the **Support**tab > Click **New User?** > Click on
the "new data
contributor form"
link

Complete Profile



03

Wait for an email response. Follow the instructions in your email.

New Data Contributor Form FAQ



Two important points to review

- The data package publisher will be your DOE Project
 Affiliation
 - ESS-DIVE is not the data package publisher
- For the PI Contact Information, please enter the primary or lead PI of your project

| DOE Proj | ect Affiliations * |
|--|--|
| associated v | Office of Science Environmental Systems Science projects/SFA programs that you are with as a team member or collaborator. Click bit.ly/ess-dive-projects to view a list of ESS ou are not affiliated with a DOE ESS project, contact ess-dive-support@lbl.gov first. |
| Your answ | er |
| | |
| | |
| Project P | rincipal Investigator (PI) Contact Information * |
| The Project the research format: PI N projects to v | rincipal Investigator (PI) Contact Information * PI (or lead PI) is the individual responsible for the preparation, conduct, and administration or grant for your affiliated project(s). Enter the Project PI contact information in the following arme, PI email. Enter PI contact information for each affiliated project. Click bit.ly/ess-dive-riew a list of ESS projects and project PIs. If your project is in this list, enter the PI contact listed there. The project PI(s) will be notified of your request to upload data to ESS-DIVE. |









How to Create and Submit Data Packages



What is a Data Package?

Data Packages have two primary components

- Data Files
- Metadata: a collection of information that describes the content and scope of the data files



Boden et al. (1999)²

Data Package: what exactly goes in it?



Author contributions

Based level of contributor effort for portions of data - author order

Data type

Particular data type from a project - e.g. continuously generated sensor data, sample data, data synthesis product



Field Campaign or Time Period

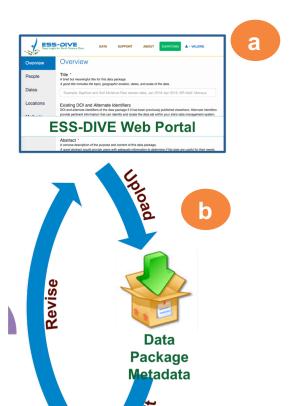
Data from a field campaign or season that need to be viewed together



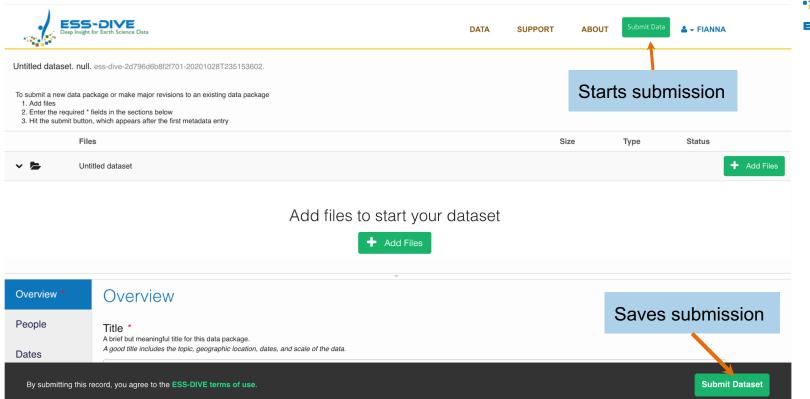
Step One: Submit

Click the Submit Data button at the top of the page to start a new data package

- Fill out all required metadata fields
- Upload data files
- Hit the Submit Dataset button at the <u>bottom</u> of the page.







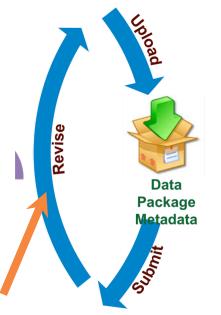
Step Two: Revise

Don't worry about completeness!

- A submitted data package is not a final product
- After you submit, you can make changes to your data package
- Submitting a data package means you are saving it in ESS-DIVE









ESS-DIVE

Worry about completeness

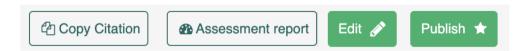
- Make sure your data files are ready for public use
- Check that your metadata is complete using "Assessment Report". More on this later!
- Hit the Publish

 button!





ESS-DIVE Data Package







| Notice: ESS-DIVE in RE | AD-ONLY mode 7/9 through 7/15 due to infrastructure upgrade. | | | | | | |
|-------------------------|--|--------------|---------|-------------|-------------|-----------|---|
| | SS-DIVE Insight for Earth Science Data | DATA | SUPPORT | ABOUT | Submit Data | ▲ + EMILY | |
| Overview * | Abstract * | | | | | | |
| People | A concise description of the purpose and content of this data package. A good abstract would provide users with adequate information to determine if the data are useful for their needs. | | | | | | |
| Dates | This is an example of an abstract. | | | | | | |
| Locations | | | | | | | |
| Methods | | | | | | | |
| | | | | | | | |
| | Manager & | | | | | | |
| | Keywords at should be associated with this data package to enable thematic searches. Search for a keyword from the list or write in your own. Tab or click enter to add to the list below with one keyword per line. The list contains GCMD |) kannumerle | | | | | |
| | Use autocomplete feature to pick from the axisting keywords. | Nayworus. | | | | | |
| | EARTH SCIENCE > AGRICULTURE > SOILS | | - | esaurus: GC | 140 | | |
| | Earth Science > Addition for the Solids | | | seaurus: GC | | × | - |
| | | | | | - | | - |
| | Data Variables | | | | 10 | 6 | - |
| | Measurement variables present in the data package. Search for a variable from the list or write in your own. Tab or click enter to add to the list below with one variable per line. The list contains GCMD is | and CF varia | bles. | | A | | |
| | Use autocomplete feature to pick from the existing variables. | | | | | 7 | |
| By submitting this reco | rd, you agree to the ESS-DIVE terms of use. | | | | | | |





- Try downloading data from https://data.ess-dive.lbl.gov
- Use ESS-DIVE's sandbox server to trying uploading dummy data https://data-sandbox.ess-dive.lbl.gov
 - This is a testing space, you don't need to register to upload data here.
 - Just update your email address first



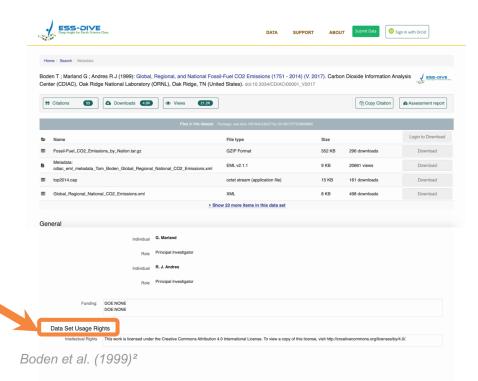
Citing ESS-DIVE Data



Citing an ESS-DIVE Data Package

Data usage rights determine citation permissions

- Check Data Set Usage Rights listed at the bottom of every Data Package
- Either Creative Commons
 Attribution (default) or Public
 Use

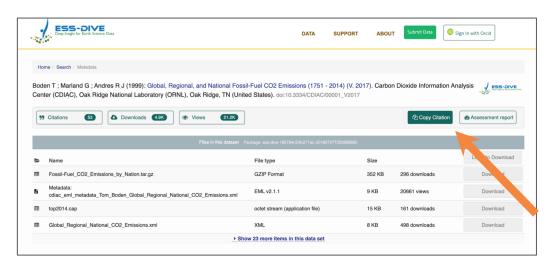




Citing an ESS-DIVE Data Package

Citing data packages is easy on ESS-DIVE

 Use Copy Citation button to store Data Package
 Citation to Clipboard



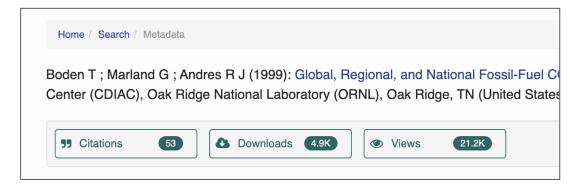
Boden et al. (1999)²

Certain metadata fields determine how the citation will look





- Every data package has citation metrics which inform you how many times your data package was cited
- Citation counts are not fully accurate because we don't get this information from journals





What happens when you hit Publish

What happens when you click *Publish*?



Confirm



01

ess-DIVE will email you to **confirm** that your publication request has been received.

Review



ESS-DIVE support team will **review** your data package and email you if they have any questions.

Public



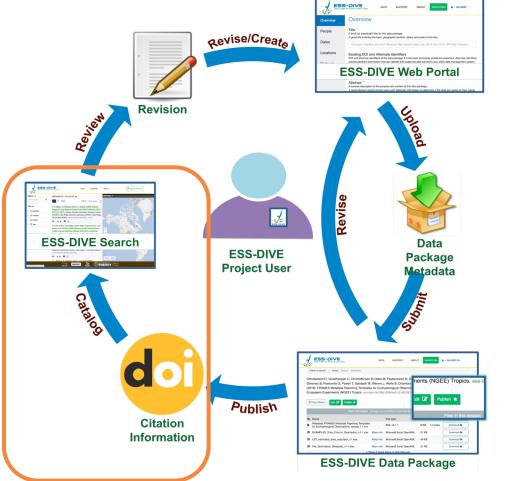
03

Once the data package has been reviewed it will be assigned a **DOI** and shared publicly.

What is a DOI?

Keeps track of data

- DOIs are persistent identifiers that are needed in order to cite your data later
- Like ESS-DIVE, most journals assign papers a DOI right when it's published





What the reviews look like from user's end





Hesham, Emily 4

EDDOI-855 ESS-DIVE Publication Request "Dissolved Inorganic Carbon and Dissolved Org...

- The submitter and dataset contact will receive a confirmation email
- ESS-DIVE team reviews data package and requests changes
- After revisions, package is made public with DOI.

Hello.

We have finished reviewing your data package and request the following improvements to your metadata before publication:

- 1. Revise your abstract to include a statement about the purpose for why these data were generated and the research question it is intended to answer. Also, include a description of the file types in your data package and any software needed to utilize them.
- 2. As with your previous data package, please include the full citations for referenced data to accompany the DOIs entered.

Note that once the data package is published, it will be publicly available for search and download. The publication action cannot be undone, however, you will still be able to edit or retire the data package. As the dataset contact, you are responsible for obtaining consent for publication from the data package authors and notifying them about its status. Please respond to this message once you are finished making these revisions so that we can continue the approval process.

Thanks,

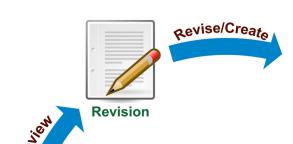
The ESS-DIVE Support Team



ESS-DIVE

After publishing, you CAN make edits to your data package

- Minor changes to your data package are allowed and encouraged
 - Fixing file headers or errors, editing abstract to include more contextual information, adjusting methods section.



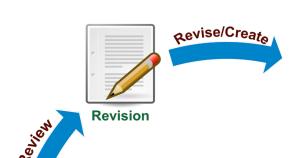




ESS-DIVE

Changing the scope of a published data package has downstream effects

- Major changes to your data package are not encouraged by ESS-DIVE
 - Changing author list, changing title name
 - Adding new data from new field campaigns







How to Check Metadata Quality

Why is it important to provide sufficient metadata? ...



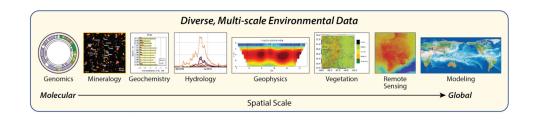
Making data **FAIR**:

Findable

Accessible

Interoperable

Reusable



Metadata can make finding data and determining data use much easier





ESS-DIVE requires metadata to enable users to understand and interpret the data. The metadata should:

- Fully explains data without need of further reference
- Makes it useful from a scientific and public perspective
- Avoids undefined acronyms and project specific terms
- Provides author contact information.
- Contains information about data generation methods

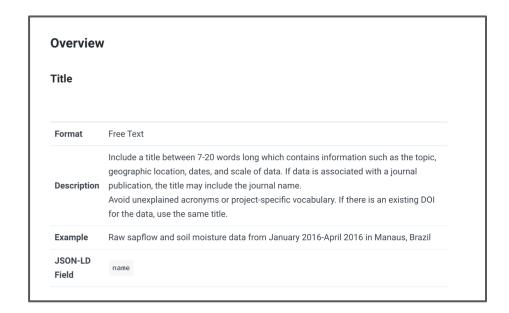




Full description of metadata guidelines found on docs.ess-dive.lbl.gov

Each metadata field is broken down by:

- Format
- Description
- Example
- JSON-LD field name for API submission







Automatic report on metadata completeness

- This report appears on private and public datasets
- There are four major categories - data packages with failed checks are required to resolve them before publication

| Metadata Assessn | nent Report |
|--|---|
| hen J; Dafflon B; Tran A; Falco N bi:10.15485/1633810 | N; Hubbard S (2020): Hybrid predictive modeling approach simulated evapotranspiration and ecosystem respiration data. Watershed Function SFA |
| ter running your metadata against our standard set | of metadata, data, and congruency checks, we have found the following potential issues. Please assist us in improving the discoverability and reusability of your research data by addressing the issues below. |
| 16 checks | |
| | Discovery: 100% complete |
| | Interpretation: 100% complete |
| | |
| Passed 13 checks out of 14 (info | rmational checks not included). |
| Warning for 1 check. Please revi | ew these warnings. |
| ▼ Failed 0 checks. | |
| | |



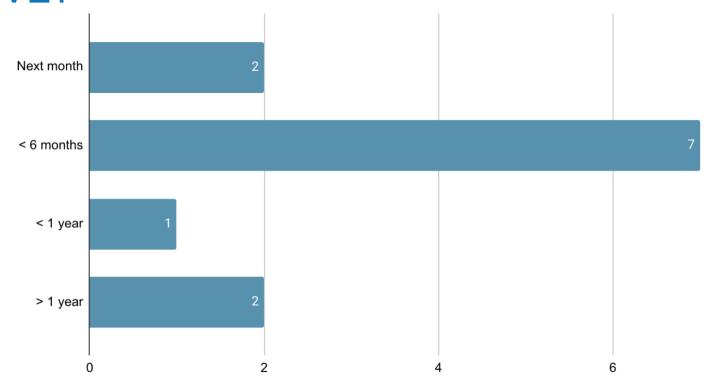


Stay tuned in to hear about advanced tools and features on ESS-DIVE in future webinars!

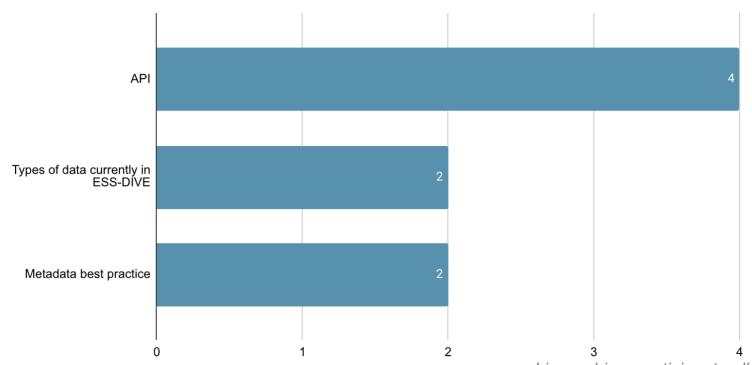
- Programmatic submissions and updates to data packages on ESS-DIVE using the Package Service API
- Customizable data portals to organize your data collections
- Community reporting formats to standardize data file organization
- An ESS-DIVE community github to review and comment on data reporting formats



When do you think you'll submit data to ESS-DIVE?



What are some topics you want more information on?





Questions?

Detailed guides are available at docs.ess-dive.lbl.gov

Tutorial videos are on **ESS-DIVE Team YouTube channel**

Contact us at ess-dive-support@lbl.gov



ESS-DIVE Glossary

List of common ESS-DIVE terms:

- DataONE The Data Observation Network for Earth (DataONE) is a
 distributed framework and sustainable cyberinfrastructure that provides open
 and secure access to Earth observational data. ESS-DIVE is a DataONE
 member.
- **DOE** The U.S. Department of Energy (DOE) is a Cabinet-level department of the United States whose mission is to ensure America's security and prosperity by addressing its energy, environmental and nuclear challenges through transformative science and technology solutions.





- DOI A Digital Object Identifier (DOI) is a unique alphanumeric string
 assigned by a registration agency (e.g., The Office of Scientific and Technical
 Information (OSTI)) to identify content and provide a persistent link to its
 location on the internet. ESS-DIVE assigns a DOI when your data package is
 published and made available electronically.
- **ESGF** The Earth System Grid Federation (ESGF) is a collaboration that develops, deploys and maintains software infrastructure for the management, dissemination, and analysis of model output and observational data. It is an interagency and international effort led by the U.S. Department of Energy with various co-funding agencies.





- ESS Environmental Systems Science (ESS) is a U.S. Department of Energy
 Office of Science program under the Biological and Environmental Research
 Program seeking to advance a robust predictive understanding of terrestrial
 surface and subsurface ecosystems.
- **ESS-DIVE** Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) is a U.S. Department of Energy repository for earth and environmental science data, models and software generated from research on terrestrial and subsurface environments.





- NCEAS The National Center for Ecological Analysis and Synthesis
 (NCEAS) at UC Santa Barbara partners with ESS-DIVE on data preservation.
 NCEAS is a DataONE member and a recognized expert in ecological data,
 digital libraries, and standards for data format.
- NERSC The National Energy Research Supercomputing Center (NERSC) is the primary scientific computing facility for the Office of Science in the Department of Energy, and is the primary data storage center for ESS-DIVE.
- ORCiD The Open Researcher and Contributor IDentifier provides anyone a
 persistent digital identifier (an ORCID iD) that distinguishes researchers from
 one another and provides a record that supports automatic links among all
 professional activities.





- OSTI The Office of Scientific and Technical Information (OSTI), a unit of the U.S. Department of Energy Office of Science, partners with ESS-DIVE on providing Digital Object Identifiers and metadata enhancement for data packages submitted to ESS-DIVE.
- Additional terms to consider defining: KBase, EMSL, JGI GOLD, NMDC,
 Dynamic data citation, Data package, Metadata, API, Text file, Comma
 separated file, NetCDF file, FAIR, ID, ISGN, Data Repository, Digital Library,
 Project Space, Data Visibility, Data Usability, Data Accessibility, JIRA,
 GitHub, Community Funds, Large/Big Data, File-level metadata, Sample-level
 metadata, 'Scalability, Resilience, and Integration', User & Preservation
 Services, Community Engagement and Partnerships

ESS-DIVE Workflow

An **ESS-DIVE** project user wishing to submit data to the repository gathers the files to be included in a data package, and uploads them via a web portal.

The data package is **shared** *publicly* and becomes available for search and download on the ESS-DIVE data portal

After the user requests the data package to be published, the ESS-**DIVE team** checks the data prior to public release, and a a systemassigned unique identifier (DOI) is assigned.



Revision

Citation Information

ESS-DIVE Search



Publish





The user specifies *metadata* associated with the data package, including author and citation information, as well as related references

The user uploads data Data **Package** *files* to be included in the Metadata data package





The user submits the data package which saves the metadata and accompanying data files to the ESS-DIVE Data Repository.

ESS-DIVE Data Package

References



- 1. Varadharajan, C., S. Cholia, C. Snavely, V. Hendrix, C. Procopiou, D. Swantek, W. J. Riley, and D. A. Agarwal (2019), Launching an accessible archive of environmental data, Eos, 100, https://doi.org/10.1029/2019EO111263. Published on 08 January 2019.
- Boden T; Marland G; Andres R J (1999): Global, Regional, and National Fossil-Fuel CO2 Emissions (1751 - 2014) (V. 2017). Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). doi:10.3334/CDIAC/00001_V2017
- 3. Icon made by Freepik from www.flaticon.com (Slide 12)
- 4. Icon made by Freepik from www.flaticon.com (Slide 12)
- 5. Icon made by Freepik from www.flaticon.com (Slide 12)

References



- 15. Icon from orcid.org (Slide 15)
- 16. Icon made by Freepik from www.flaticon.com (Slide 15)
- 17. Icon made by Tomas Knop from www.flaticon.com (Slide 15)
- 18. Icon made by Freepik from www.flaticon.com (Slide 20)
- 19. Icon made by Freepik from www.flaticon.com (Slide 20)
- 20. Icon made by Justicon from www.flaticon.com (Slide 20)
- 21. Icon made by Eucalyp from www.flaticon.com (Slide 20)





- 22. Icon made by Freepik from www.flaticon.com (Slide 32)
- 23. Icon made by Freepik from www.flaticon.com (Slide 32)
- 24. Icon made by Freepik from www.flaticon.com (Slide 32)